



DEUTSCHES  
PATENTAMT

21 Aktenzeichen: 195 31 222.8  
22 Anmeldetag: 24. 8. 95  
43 Offenlegungstag: 27. 2. 97

71 Anmelder:  
Siemens AG, 80333 München, DE

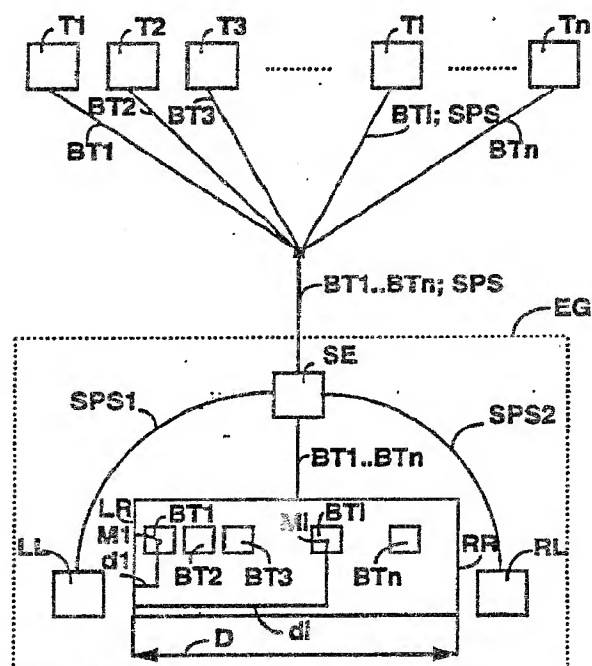
72 Erfinder:  
Stier, Sebastian, Dipl.-Phys. Dr., 85653 Aying, DE

56 Entgegenhaltungen:  
DE 44 20 212 A1  
EP 01 53 903 A1  
Flanagan, J.L.: »Speech Technology and Computing:  
A Unique Partnership«, in: IEEE Communications  
Magazine, May 1992, S. 84-89;

Prüfungsantrag gem. § 44 PatG ist gestellt

54 Verfahren zur Zuordnung eines Sprachsignals eines Teilnehmers einer Multipoint-Videokonferenz zu einem Bildbereich in einem Bildschirm

57 Verfahren zur Zuordnung eines Sprachsignals (SPS) eines Teilnehmers einer Multipoint-Videokonferenz (MV) zu einem Bildbereich in einem Bildschirm (BS).  
Die Erfindung betrifft ein Verfahren, bei dem das Sprachsignal (SPS) auf einen ersten Teil (SPS1) und einen zweiten Teil (SPS2) aufgeteilt wird. Die Aufteilung erfolgt abhängig von der Position eines Bildes (BTi) eines Teilnehmers (Ti), dem das Sprachsignal (SPS) zugeordnet wird, innerhalb eines Bildschirms (BS).  
Hierdurch wird ein räumlicher Klangeffekt erzielt, da mit dem ersten Teil (SPS1) ein erster Lautsprecher (LL) angesteuert wird, und mit dem zweiten Teil (SPS2) ein zweiter Lautsprecher (RL) angesteuert wird. Der erste Lautsprecher (LL) ist links neben dem Bildschirm (BS) angeordnet, und der zweite Lautsprecher (RL) rechts neben dem Bildschirm (BS).  
Der räumliche Klangeffekt vermittelt dem Betrachter des Bildschirms (BS) einen räumlichen Eindruck und ermöglicht es ihm so, das Sprachsignal intuitiv der "richtigen" Position innerhalb des Bildschirms (BS) zuzuordnen.



Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Die Erfindung betrifft ein Verfahren, das bei Videokonferenzen mit mehr als zwei Teilnehmern, sog. Multipoint-Videokonferenzen zur Anwendung kommt. Bei Videokonferenzen, insbesondere bei Multipoint-Videokonferenzen sieht ein Teilnehmer seine Partner in verschiedenen Fenstern, also verschiedenen Bildbereichen, die auf einem Bildschirm dargestellt werden.

Der Gesprächsablauf der Multipoint-Videokonferenz muß entweder von einem Moderator koordiniert werden, oder es muß genaue Regeln dafür geben, wer das Wort erhält. Somit wird bei einer Multipoint-Videokonferenz nur jeweils ein Sprachsignal genau eines Teilnehmers übertragen, da nur ein Teilnehmer zu jedem Zeitpunkt "Rederecht" erhält.

Dies ist schon deswegen von Bedeutung, weil ein Teilnehmer nicht ausreichend schnell und einfach erkennt, wer gerade redet, also "Rederecht" hat, d. h. in welchem Fenster, also in welchem Bildbereich des Bildschirms gerade "geredet" wird. Es wird außerdem äußerst mühsam, den Gesprächsablauf zu verfolgen, wenn mehrere Partner kurz hintereinander für nur sehr kurze Zeit sprechen, d. h. wenn das "Rederecht" häufig wechselt. In natürlichen Gesprächssituationen tritt dies sehr häufig auf und bereit keine Schwierigkeiten. Die Gesprächssituation in Multipoint-Videokonferenzen jedoch erscheint durch die Einschränkung, daß zu jedem Zeitpunkt immer ein Teilnehmer "Rederecht" besitzt, künstlich und eingeschränkt.

Der Erfindung liegt das Problem zugrunde, ein Verfahren anzugeben, durch das ein Sprachsignal eines Teilnehmers einer Multipoint-Videokonferenz zu einem Bildbereich eines Bildschirms, in dem der entsprechende Teilnehmer dargestellt wird, zugeordnet wird.

Das Problem wird durch das Verfahren gemäß Patentanspruch 1 gelöst.

Durch die Ansteuerung zweier Lautsprecher, wobei jeweils ein Lautsprecher links und ein Lautsprecher rechts von dem Bildschirm angeordnet ist, mit jeweils einem Teil des gesamten Sprachsignals, der abhängig ist vom Abstand des Bildbereiches, in dem der jeweils sprechende Teilnehmer dargestellt ist, von den Rändern des Bildschirms, wird eine intuitive, leichtere Zuordnung des Sprachsignals zu dem jeweils geradeprechenden Teilnehmer der Multipoint-Videokonferenz für den Betrachter des Bildschirms erreicht.

In den Figuren ist ein bevorzugtes Ausführungsbeispiel beschrieben.

Es zeigen

Fig. 1 eine Skizze, in der eine Anordnung dargestellt ist, in der das erfindungsgemäße Verfahren durchgeführt wird; außerdem sind die einzelnen Bild- und Sprachsignale und die Aufteilung des Sprachsignals skizziert;

Fig. 2 ein Ablaufdiagramm, in dem das erfindungsgemäße Verfahren dargestellt ist.

Anhand der Fig. 1 und 2 wird die Erfindung weiter erläutert.

In Fig. 1 ist eine Anordnung skizziert, in der eine Multipoint-Videokonferenz MV mit einer Anzahl  $n$  von Teilnehmern  $T_i$  der Multipoint-Videokonferenz MV beschrieben ist. Ein Index  $i$  ist eine natürliche Zahl im Bereich von 1 bis  $n+1$ , und bezeichnet jeden Teilnehmer  $T_i$  der Multipoint-Videokonferenz MV eindeutig.

Für einen  $n+1$ -ten Teilnehmer, also für den Teilnehmer  $T_{n+1}$  ist ein Endgerät EG dargestellt.

Bei der Multipoint-Videokonferenz MV wird von je-

dem Teilnehmer  $T_i$  jeweils ein Bild  $BT_i$  an dem  $n+1$ -ten Teilnehmer  $T_{n+1}$  übertragen.

Die Bilder  $BT_i$  sind jeweils die Videodaten der Videokonferenz, die jeweils von jedem Teilnehmer  $T_i$  an den  $n+1$ -ten Teilnehmer  $T_{n+1}$  der Multipoint-Videokonferenz MV gesendet werden. Da bei der Multipoint-Videokonferenz MV jeweils nur ein Teilnehmer "Rederecht" besitzt, d. h. nur ein Teilnehmer  $T_i$  hat zu einer bestimmten Zeit das ausschließliche Recht und die alleinige Möglichkeit, ein Sprachsignal SPS zu senden, wird immer nur ein Sprachsignal SPS, das genau dem Teilnehmer  $T_i$  zugeordnet wird, der zu dem bestimmten Zeitpunkt das "Rederecht" besitzt, an das Endgerät EG des  $n+1$ -ten Teilnehmer gesendet.

Diese Regelung zur Verteilung des "Rederechts" ist bei einer Multipoint-Videokonferenz MV nötig, um ein Chaos im Rahmen der Multipoint-Videokonferenz MV zu vermeiden.

Wenn der  $n+1$ -te Teilnehmer  $T_{n+1}$  zu einer bestimmten Zeit kein "Rederecht" besitzt, so werden also von  $n$  Teilnehmern genau  $n$  Bilder  $BT_i$  der  $n$  Teilnehmer  $T_i$  und ein Sprachsignal SPS des Teilnehmers, der zu dem Zeitpunkt gerade "Rederecht" besitzt, an den  $n+1$ -ten Teilnehmer  $T_{n+1}$  gesendet.

Dieses gesamte Signal, also  $n$  Bilder  $BT_1 \dots BT_n$  und das Sprachsignal SPS werden von dem Endgerät EG des  $n+1$ -ten Teilnehmer  $T_{n+1}$  empfangen und gespeichert 1.

In einer Schaltereinheit SE werden die  $n$  Bilder  $BT_1 \dots BT_n$  der  $n$  Teilnehmer  $T_i$  auf einem Bildschirm BS dargestellt, wobei die Bilder  $BT_i$  nebeneinander bzw. teilweise überlappend auf dem Bildschirm BS dargestellt werden 2.

Das Sprachsignal SPS wird in der Schaltereinheit SE in einen ersten Teil SPS1 des Sprachsignals SPS und in einen zweiten Teil SPS2 des Sprachsignals SPS aufgeteilt.

In der Schaltereinheit SE ist es bekannt, zu welchem Teilnehmer  $T_i$  das Sprachsignal SPS gehört, d. h. es ist bekannt, welcher Teilnehmer  $T_i$  der Multipoint-Videokonferenz MV zu jedem Zeitpunkt "Rederecht" besitzt.

Durch Aufteilung des Sprachsignals SPS in den ersten Teil SPS1 des Sprachsignals SPS und den zweiten Teil SPS2 des Sprachsignals SPS wird der Effekt erzielt, daß für den Betrachter und Zuhörer, d. h. für den  $n+1$ -ten Teilnehmer  $T_{n+1}$  intuitiv erkennbar ist anhand des durch die Teile des Sprachsignals SPS hervorgerufenen Audioeffekts, welcher Teilnehmer  $T_i$ , der ja in Form eines Bildes  $BT_i$  auf dem Bildschirm BS dargestellt ist, gerade "Rederecht" besitzt.

Die Bilder  $BT_i$  können sowohl ein Standbild als auch eine Folge von Bildern, also eine Aufnahme einer Videokamera, wie das bei einer Multipoint-Videokonferenz häufig üblich ist, sein.

Mit dem ersten Teil SPS1 des Sprachsignals SPS wird ein erster Lautsprecher LL angesteuert 3.

Mit dem zweiten Teil SPS2 des Sprachsignals SPS wird ein zweiter Lautsprecher RL angesteuert 3 (vgl. Fig. 2).

Hierbei ist der erste Lautsprecher LL links von dem Bildschirm BS angeordnet und der zweite Lautsprecher RL ist rechts neben dem Bildschirm BS angeordnet.

Die Aufteilung des Sprachsignals SPS in den ersten Teil SPS1 des Sprachsignals SPS und in den zweiten Teil SPS2 des Sprachsignals SPS erfolgt in der im weiteren beschriebenen Weise.

Wenn z. B. der erste Teilnehmer  $T_1$  "Rederecht" erhält, ist das Sprachsignal dem ersten Teilnehmer  $T_1$  zu-

geordnet. Dies ist der Schaltereinheit SE bekannt. In der Schaltereinheit SE wird die Position des Bildes BT1 innerhalb des Bildschirms BS ermittelt.

Dabei wird der horizontale Abstand  $d_1$  des ersten Bildes BT1 von einem linken Rand LR des Bildschirms BS berechnet. Als Referenzpunkt zur Bestimmung des Abstandes des Bildes BT1 von dem linken Rand LR des Bildschirms BS kann ein beliebiger Punkt innerhalb oder am Rand des Bildes BT1 verwendet werden. Üblicherweise bietet sich ein Mittelpunkt M1 des Bildes BT1 als Referenzpunkt zur Bestimmung des Abstandes  $d_1$  vom linken LR des Bildschirms BS an.

Allgemein bedeutet die im vorigen beschriebene Vorgehensweise, daß, wenn der i-te Teilnehmer  $T_i$  "Rederecht" erhält, der Abstand  $d_i$  des i-ten Bildes  $BT_i$  von dem linken Rand LR des Bildschirms BS ermittelt wird, wobei beispielsweise als Referenzpunkt der Mittelpunkt  $M_i$  des i-ten Bildes  $BT_i$  verwendet wird.

Nun wird das Verhältnis des Abstandes  $d_i$  des Bildes  $BT_i$  des Teilnehmers  $T_i$ , der gerade "Rederecht" hat, durch eine Gesamtbreite D des Bildschirms BS dividiert.

Das Verhältnis  $\frac{d_i}{D}$  bildet einen Faktor, mit dem das Sprachsignal SPS, also die Lautstärke des Sprachsignals SPS, multipliziert wird, und das Ergebnis bildet den ersten Teil SPS1 des Sprachsignals SPS, mit dem der erste Lautsprecher LL angesteuert wird. Der erste Teil SPS1 des Sprachsignals SPS ist also direkt abhängig von der Position des Bildes  $BT_i$  des Teilnehmers  $T_i$ , zu dem das Sprachsignal SPS zugeordnet ist, d. h. des Teilnehmers  $T_i$ , der gerade "Rederecht" besitzt.

Der Rest des Sprachsignals SPS, also 1-SPS1, bildet den zweiten Teil des Sprachsignals SPS2.

Anders ausgedrückt ergibt sich der zweite Teil SPS2 aus folgendem Zusammenhang:

$$(1 - d_i/D) \cdot \text{SPS}.$$

Somit ist der zweite Teil SPS2 des Sprachsignals SPS direkt abhängig von dem Abstand des Bildes  $BT_i$  des Teilnehmers  $T_i$ , der gerade "Rederecht" hat, von einem rechten Rand RR des Bildschirms BS.

Durch die Aufteilung der Gesamtlautstärke des Sprachsignals SPS in den ersten Teil SPS1 und in den zweiten Teil SPS2 wird ein räumlicher Klangeffekt für den  $n+1$ -ten Teilnehmer  $T_{n+1}$ , also den Betrachter des Bildschirms BS erreicht, da für den Betrachter und gleichzeitig Zuhörer der Multipoint-Videokonferenz MV die Position des Bildes  $BT_i$  des Teilnehmers  $T_i$  innerhalb des Bildschirms BS intuitiv erschlossen werden kann aus dem räumlichen Klangeindruck.

Da der Betrachter und Zuhörer der Multipoint-Videokonferenz MV aus dem Klang, also aus der Aufteilung des Sprachsignals SPS in den ersten Teil SPS1 des Sprachsignals SPS und den zweiten Teil SPS2 des Sprachsignals SPS auf die Position des Bildes  $BT_i$  des Teilnehmers  $T_i$ , der "Rederecht" besitzt, innerhalb des Bildschirms BS schließen kann, ist es für den Betrachter also nur aufgrund des Sprachsignals SPS bzw. des ersten Teils SPS1 des Sprachsignals SPS und des zweiten Teils SPS2 des Sprachsignals SPS möglich herauszufinden, welcher Teilnehmer  $T_i$  zu einem Zeitpunkt "Rederecht" besitzt.

Dies erleichtert einem Teilnehmer das Verfolgen einer Multipoint-Videokonferenz MV erheblich, da eine lästige "Suche" des gerade sprechenden Teilnehmers  $T_i$  auf dem Bildschirm BS, entfällt.

#### Patentansprüche

1. Verfahren zur Zuordnung eines Sprachsignals (SPS) eines Teilnehmers einer Multipoint-Videokonferenz (MV) zu einem Bildbereich eines Bildschirms (BS),

- bei dem in einem Endgerät (EG) ein Sprachsignal (SPS) eines Teilnehmers der Multipoint-Videokonferenz (MV) empfangen wird,
- bei dem in dem Endgerät (EG) Bilder ( $BT_i$ ;  $i=1 \dots n$ ) aller Teilnehmer ( $T_i$ ) der Multipoint-Videokonferenz (MV) empfangen werden,
- bei dem die Bilder ( $BT_i$ ) der Teilnehmer ( $T_i$ ) der Multipoint-Videokonferenz (MV) auf dem Bildschirm (BS) nebeneinander oder teilweise überlappend dargestellt werden,
- bei dem ein links neben einem linken Rand (LR) des Bildschirms (BS) positionierter erster Lautsprecher (LL) mit einem ersten Teil (SPS1) des Sprachsignals (SPS) angesteuert wird, dessen Anteil an dem Sprachsignal (SPS) der relativen horizontalen Position des Bildes ( $BT_i$ ) des Teilnehmers ( $T_i$ ) der Multipoint-Videokonferenz (MV) bezüglich des linken Randes (LR) des Bildschirms (BS) entspricht, und
- bei dem ein rechts neben einem rechten Rand (RR) des Bildschirms (BS) positionierter zweiter Lautsprecher (RR) mit einem zweiten Teil (SPS2) des Sprachsignals (SPS) angesteuert wird, dessen Anteil an dem Sprachsignal (SPS) der relativen horizontalen Position des Bildes ( $BT_i$ ) des Teilnehmers ( $T_i$ ) der Multipoint-Videokonferenz (MV) bezüglich des rechten Randes (RR) des Bildschirms (BS) entspricht.

Hierzu 2 Seite(n) Zeichnungen

- Leerseite -

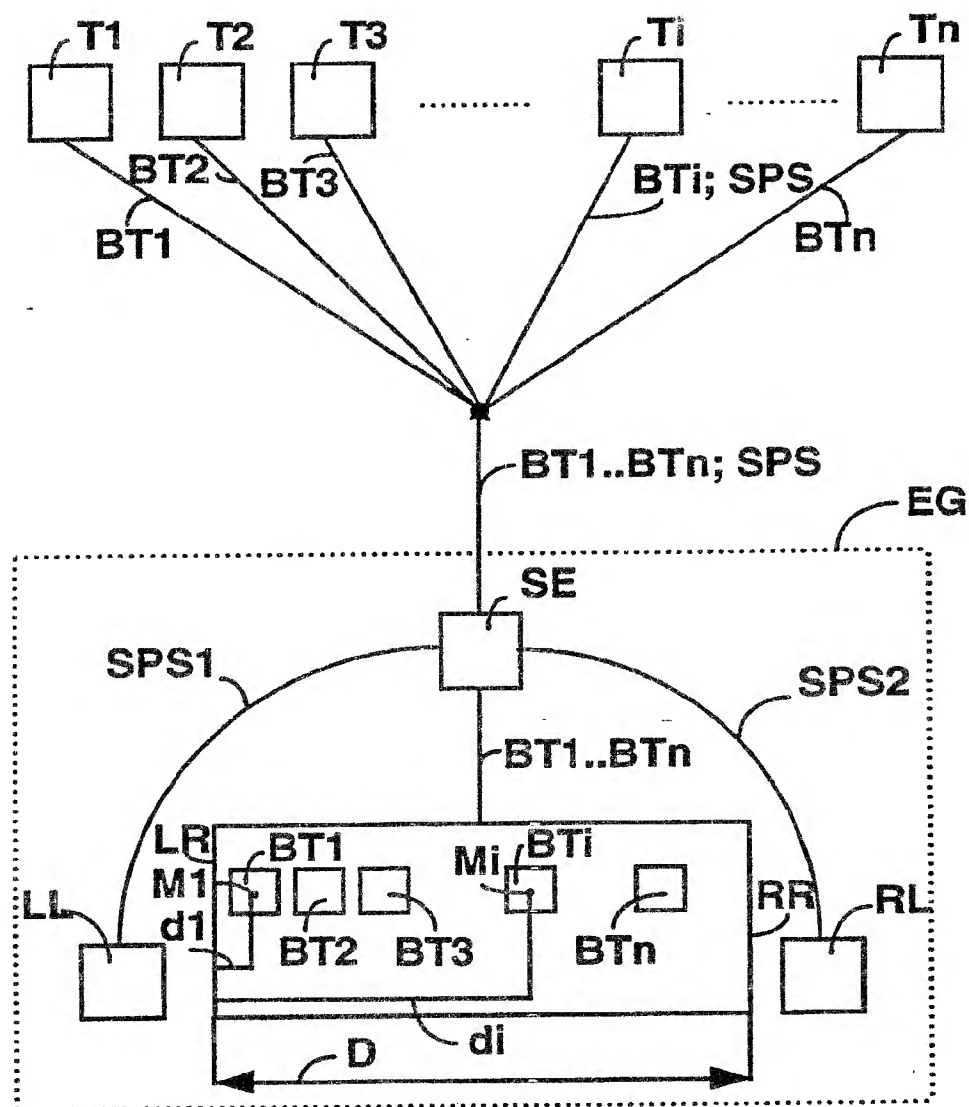
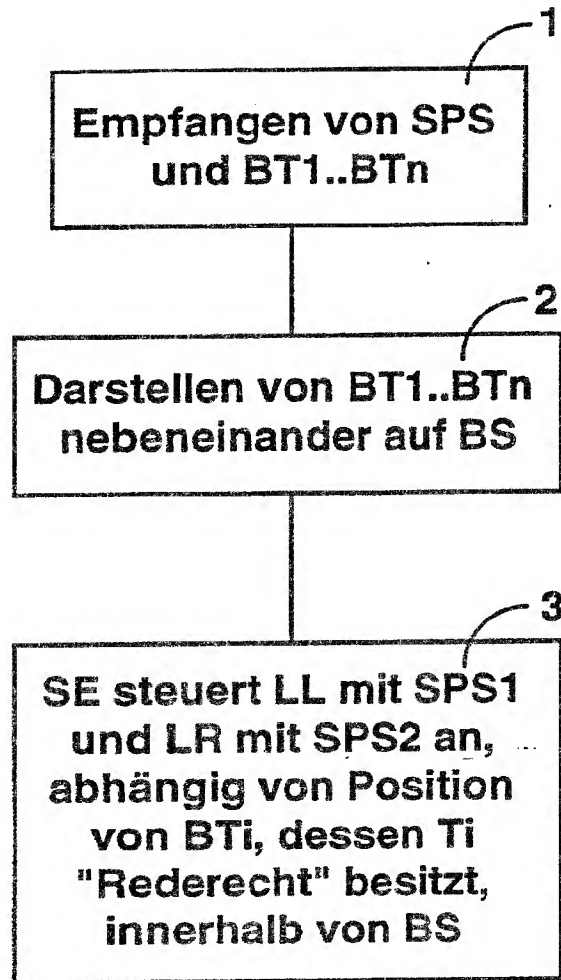


Fig. 1



**Fig. 2**